

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

Wittgenstein’s ‘Notorious’ Paragraph About the Gödel Theorem:  
Recent Discussions<sup>1</sup>

Juliet Floyd and Hilary Putnam

How would it be if *Principia Mathematica* (hereafter “PM”) turned out to be  $\omega$ -inconsistent?<sup>2</sup> In our brief “Note on Wittgenstein’s ‘Notorious Paragraph’ About the Gödel Theorem” (2000)<sup>3</sup> we showed that in 1937, when he wrote down the most ‘notorious’ of his remarks about Gödel, Wittgenstein may well have been pondering this question.<sup>4</sup> Viewed in this light, we argued, his remarks may be seen to contain the germ of a significant philosophical insight about Gödel's theorem, rather than a hopeless effort to refute it. The purpose of our "Note" was to characterize this insight, detaching it from the disputed question whether Wittgenstein fully understood the Gödel theorem.

The insight, as we construed it, asks us to appreciate the philosophical naiveté and/or unclarity<sup>5</sup> involved in taking the following claim to be a straightforward truth:

*Claim:*

Gödel’s theorem *proves* that

- 1) there is a well-defined notion of “mathematical truth” applicable to every formula of *Principia Mathematica*

and that

- 2) if *Principia Mathematica* is consistent, then some “mathematical truths” in the sense of 1) above are undecidable in PM.

---

<sup>1</sup> This is a revised version of our essay “Bays, Steiner and Wittgenstein’s ‘Notorious’ Paragraph About the Gödel Theorem”, *The Journal of Philosophy* 103, 2 (February 2006): 101-110, incorporating changes and additions in light of responses to the original publication communicated by Timothy Bays as well as elaborations included in Putnam’s “A Note on Steiner on Wittgenstein, Gödel and Tarski”, *Iyyun* 57 (January 2008): 83-93. JF is grateful to Warren Goldfarb for helpful comments on a penultimate draft of this paper.

<sup>2</sup> A formal system  $L$  is  $\omega$ -inconsistent if there exists some well-formed formula  $P(v)$ , expressing a predicate of natural numbers and with no free variable other than  $v$ , such that  $(\exists v)Pv$  is provable in  $L$  and yet so are all the formulas  $\neg P(\underline{0}), \neg P(\underline{1}), \neg P(\underline{2}), \dots$  (where  $\underline{0}, \underline{1}, \underline{2}, \dots$  are the formal expressions for the natural numbers in  $L$ ).  $L$  is  $\omega$ -consistent if there is no such well-formed formula.

<sup>3</sup> Juliet Floyd and Hilary Putnam, “A Note on Wittgenstein’s ‘Notorious Paragraph’ About the Gödel Theorem”, *The Journal of Philosophy* XCVII, 11 (November 2000): 624-632.

<sup>4</sup> The “notorious” remarks, written in the fall of 1937, are published in Ludwig Wittgenstein, *Remarks on the Foundations of Mathematics*, revised edition, (Cambridge, MA: MIT Press, 1978) Part I Appendix III §8.

<sup>5</sup> We labeled this unclarity “metaphysical”, Floyd and Putnam, p. 632.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

This insight had never been specifically associated with Wittgenstein’s ‘notorious’ remarks, and does not leap to the eye on their surface. Moreover, there are rival interpretations of the remarks that commit their authors to the view that Wittgenstein *could* not have been thinking along such lines, and/or to the view that if he was so thinking, then he was mistaken.

In what follows we explain why we chose to extract and emphasize this particular insight from the notorious remarks. Our treatment of Wittgenstein takes place within a rather complicated context of contemporary debate about the philosophical significance of Wittgenstein’s writings on Gödel; another aim of the essay is to survey this terrain.

We do not claim that our reading exhausts the themes at stake in Wittgenstein’s remarks about the Gödel theorem, nor that it offers a general approach to Wittgenstein’s remarks on mathematics as a whole, which were after all written down tentatively for his own use, and not intended for publication. We do wish to defend the force and interest of our interpretation, however, on the ground that it does more credit, not only to Wittgenstein’s philosophical perspicacity, but to the interest of the paragraph itself. Certainly we think it important to stress that Wittgenstein’s remarks, and the insight we see contained in them, are not at all an attempt to refute Gödel. Moreover, insofar as they may be seen to broach the insight we describe above, the remarks are of interest to philosophy of mathematics today.

First, we review the claims that were forwarded in our “Note”. Second, we consider a rival interpretation of Wittgenstein’s remarks offered by Mark Steiner in his essay “Wittgenstein as His Own Worst Enemy: The Case of Gödel’s Theorem” (2001).<sup>6</sup> Steiner’s essay, written partly in response to earlier work of Floyd’s (1995)<sup>7</sup>, was composed before he knew of our interpretation, so here we will be laying out explicitly where we take our differences with Steiner to lie, and offering a response, although a full consideration of Steiner’s interpretation of the notorious passage lies outside the scope of this essay. Steiner has said in conversation that our “Note” would not have led him to withdraw his criticism of Wittgenstein’s remarks on the Gödel Theorem, but we hope that

---

<sup>6</sup> *Philosophia Mathematica* IX (2001): 257-279.

<sup>7</sup> Juliet Floyd, "On Saying What You Really Want to Say: Wittgenstein, Gödel and the Trisection of the Angle" in ed. J. Hintikka, *From Dedekind to Gödel: The Foundations of Mathematics in the Early Twentieth Century*, *Synthese Library Vol. 251* (Kluwer Academic Publishers, 1995), pp. 373-426.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

what we write here may help to better sort out the underlying issues. In our third and concluding section we consider the implications of our objections to Steiner for some recent criticisms of our “Note” authored by Timothy Bays. Bays alleged that we have given a philosophical answer to the question “How would it be if PM turned out to be  $\omega$ -inconsistent?” that is “almost certainly false” and based upon an argument that is “inadequate”.<sup>8</sup> We defend our reading against his allegations here. While Bays cites Steiner’s essay and appears to be indebted to parts of Steiner’s interpretation of Wittgenstein, he departs from Steiner in significant philosophical ways, as he has come to stress himself in subsequent (as yet unpublished) remarks.<sup>9</sup> In particular, he forwards a view of the notion of truth that we find problematic, both in application to Wittgenstein and in application to the insight we extracted from his ‘notorious’ remarks.

Some may feel that the most we can hope to show is that other readers have been uncharitable to Wittgenstein, and this would not by itself suffice to show that they are *wrong* about him. But we believe that, other things being equal, charitable interpretations should be preferred to uncharitable ones, especially when one is dealing with a great philosopher. This of course is not to say that one should always interpret a philosopher so that he or she comes out *right*: we are not Wittgensteinian “fundamentalists”. We agree, for example, that Wittgenstein’s rejectionist attitude to set theory was mistaken.<sup>10</sup> This, however, is a separate issue so far as the argument of our “Note” is concerned.

### 1. *What Did We Actually Claim?*

---

<sup>8</sup> Timothy Bays, “On Floyd and Putnam on Wittgenstein on Gödel”, *The Journal of Philosophy* CI,4 (April 2004): 197-210.

<sup>9</sup> Bays has posted a manuscript replying to the earlier version of this essay on his website at (<http://www.nd.edu/~tbays/papers/index.html>) titled “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.” So the debate is ongoing. Though there is much to say, we will not fully reply to this in-progress work here. We shall, however, separate out his reply from Steiner’s interpretation, for Bays has retracted the footnote that led us to think he was relying on Steiner in the originally published version of this essay (see Bays 2004 208 n. 27). See footnote 19 below.

<sup>10</sup> Steiner discusses this rejection of set theory in his “Wittgenstein as His Own Worst Enemy”; Putnam glosses the form it takes in RFM II in “Wittgenstein and the Real Numbers”, in *Wittgenstein and the Moral Life: Essays in Honor of Cora Diamond*, ed. Alice Crary (Cambridge, MA: MIT Press, 2007), pp. 235-250; compare the closing paragraph of William Tait, “Wittgenstein and the “Skeptical Paradoxes”, *The Journal of Philosophy* 83 (1986): 475-488, reprinted in *The Provenance of Pure Reason: Essays in the Philosophy of Mathematics and Its History* (New York: Oxford University Press, 2005), pp. 198-211.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

Here are the ‘notorious’ remarks on Gödel, consisting of two paragraphs published subsequently as *Remarks on the Foundations of Mathematics I*, appendix III, §8:

I imagine someone asking my advice; he says: “I have constructed a proposition (I will use ‘*P*’ to designate it) in Russell’s symbolism, and by means of certain definitions and transformations it can be so interpreted that it says: ‘*P* is not provable in Russell’s system.’ Must I not say that this proposition on the one hand is true, and on the other hand is unprovable? For suppose it were false; then it is true that it is provable. And that surely cannot be! And if it is proved, then it is proved that it is not provable. Thus it can only be true, but unprovable.”

Just as we ask, “‘Provable’ in what system?,” so we must also ask, “‘True’ in what system?” “True in Russell’s system” means, as was said, proved in Russell’s system, and “false in Russell’s system” means the opposite has been proved in Russell’s system.--Now what does your “suppose it is false” mean? *In the Russell sense* it means, “suppose the opposite is proved in Russell’s system”; *if that is your assumption* you will now presumably give up the interpretation that it is unprovable. And by “this interpretation” I understand the translation into this English sentence. - If you assume that the proposition is provable in Russell’s system, that means it is true *in the Russell sense*, and the interpretation “*P* is not provable” again has to be given up. If you assume that the proposition is true in the Russell sense, *the same* thing follows. Further: if the proposition is supposed to be false in some other than the Russell sense, then it does not contradict this for it to be proved in Russell’s system. (What is called “losing” in chess may constitute winning in another game.)

Inspired by these remarks, we argued as follows. Suppose that to our surprise we have discovered a proof of the negation of a Gödel sentence, “ $\neg P$ ”, in *Principia Mathematica*.<sup>11</sup> Suppose too that PM is consistent. Under these suppositions it is a well-known, uncontroversial consequence of Gödel’s theorem that PM will be  $\omega$ -inconsistent. As may be seen from inspection of the definition of “ $\omega$ -inconsistent” (see footnote 2), this means that every model of PM must contain entities that are not natural numbers. In fact, in any such model *every* numerical predicate with an infinite extension will “overspill”, that is, contain some elements that are not natural numbers. But then, because our original rigorization of the syntactic notions applied in the English sentence

---

<sup>11</sup> P has the form:  $\neg(\exists x)(\text{NaturalNo.}(x).\text{Proof}(x,t))$ , where “t” abbreviates a numerical expression whose value calculates out to be the Gödel number of P itself, “Proof” abbreviates a predicate which is supposed to define an effectively calculable relation which holds between two natural numbers n,m just in case n is the Gödel number of a proof whose last line is the formula with Gödel number m, and “NaturalNo.(x)” is the predicate of PM we interpret as “x is a natural number”.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

came through Gödel numbering with the natural numbers alone (as in Gödel’s original (1931) paper<sup>12</sup>), our initial way of translating into English or German “P” (the formula of PM whose proposed translation was “P is not provable in PM”) would, in this context, become more nuanced and complicated, would have to be “given up” in its original form, just as Wittgenstein observed.

Neither we nor Wittgenstein as we understand him were trying to argue that there are *no* circumstances under which we may legitimately translate “P” into the English sentence “‘P’ is not provable” (or say that P itself is “true but unprovable”).<sup>13</sup> Indeed, our interpretation prescind altogether from talk about all proper or possible translations or interpretations of P.<sup>14</sup> Instead, we take Wittgenstein to have been exploring the context-dependent *sense* of the claim that “there are true but unprovable sentences of arithmetic” as it applies both to PM and to our natural language, and doing so within a particular philosophical and historical situation. We also connect Wittgenstein’s remark about “giving up” an interpretation of “P” with his (commendable) rejection of a notion of *interpretation* which in principle wholly depends upon formalized mathematical language: as we wrote, Wittgenstein was “denying that a formal system *could* provide us with a standard of truth or clarity that is, in principle, inaccessible to a natural language” (p. 632). Of course, the argument in our “Note” itself then turns on a partly informal understanding of the notion a sentence’s being “true in an interpretation” (or “satisfied in a model”). As all agree, in 1937 model theory was not yet an established branch of mathematics, and Wittgenstein may be forgiven for leaving its development out of his purview. Moreover, we believe that the later development of formal semantics and model theory fail to impugn the importance of (what we are calling) Wittgenstein’s philosophical insight about the Gödel theorem.

## 2. Steiner’s Interpretation of the Notorious Remarks

Steiner’s writing is unfailingly interesting, and the essay in which he claimed that in the notorious remarks Wittgenstein “slips into trying to refute the [Gödel

---

<sup>12</sup> *Kurt Gödel Collected Works, Vol. I*, eds. S. Feferman et.al., (New York: Oxford University Press, 1986), pp. 145-195.

<sup>13</sup> Despite what Bays says (2004, pp. 198, 201, 202-3, 208n).

<sup>14</sup> Again, unlike Bays's interpretation, which we shall discuss below.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

incompleteness] theorem”, is certainly no exception.<sup>15</sup> We shall, however, offer reasons to doubt that Steiner’s interpretation offers a satisfactory interpretation of the notorious remarks.<sup>16</sup>

According to Steiner’s interpretation, Wittgenstein aspired in the notorious remarks to show that the claim that “there are true but unprovable sentences of arithmetic” is false, nothing but a misguided philosophical interpretation of the theorem. It is the last two sentences of the second paragraph (starting with the words “Further: if the proposition is supposed to be false in some other than the Russell sense”) that Steiner describes as “[Wittgenstein’s] ‘refutation’ of Gödel’s theorem” (Steiner 2001, 274). As Steiner reads the section, up to the words “Thus it can only be true but unprovable”, Wittgenstein is describing what he takes to be Gödel’s proof, and the words *Further: if the proposition is supposed to be false in some other than the Russell sense, then it does not contradict this for it to be proved in Russell’s system. (What is called “losing” in chess may constitute winning in another game.)*” are supposed to refute that proof. As Steiner summarizes his reading of these sentences (Steiner 2001, p. 261):

The refutation goes (I abbreviate): there is no contradiction in a false, but provable sentence - what is false is context (or ‘game’) dependent. The very same words might sometimes express a truth and sometimes a falsehood. Thus Gödel’s proof rests on an elementary mistake.

This does look at first blush like a *possible* reading of what is going on in the Gödel remarks, and squares with a long history of reading the passage as evincing Wittgenstein’s supposed view that “true in *Principia Mathematica*” and “provable in *Principia Mathematica*” amount under philosophical analysis to the same thing.<sup>17</sup> The purpose of our “Note” was, however, to broach and defend a different reading, one which did not presuppose this reduction of the notion of *truth* to that of *proof* or *game*.

---

<sup>15</sup> Mark Steiner, “Wittgenstein as his Own Worst Enemy: The Case of Gödel’s Theorem”, *Philosophia Mathematica* 9 (2001): 257-279.

<sup>16</sup> A full survey of the issues raised by Steiner’s own rather sophisticated understanding of Wittgenstein on Gödel lies outside the scope of this reply, but some of them have been discussed in Floyd, “Prose versus Proof: Wittgenstein on Gödel, Tarski and Truth” *Philosophia Mathematica* 3, 9 (2001): 901-928.

<sup>17</sup> This is discussed in Floyd’s “On Saying What You Really Want to Say: Wittgenstein, Gödel and the Trisection of the Angle”, pp. 376-7, note 21.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

Steiner maintains that the argument he attributes to Wittgenstein in the notorious paragraph must be wrong. For, as he explains, as a result of Tarski’s analysis of the notion of a sentence being *true in a formalized language* (or *satisfied in a structure*) we can give a proof in formal semantics of the Gödel theorem that demonstrates the existence of “true but unprovable” sentences of arithmetic. He writes,

...we can use Tarski’s concept of truth to prove the Gödel theorem directly. It is a *mathematical theorem* (of set theory) that the Gödel sentence  $P$  is true in Tarski’s sense if and only if not provable in PA. [Peano Arithmetic]. (That is, Gödel constructs the sentence  $P$  to have this property). Then if  $P$  is false, it is provable in PA, and so we have a false theorem of PA, which is impossible, because we have another *mathematical theorem* that all theorems of PA are true (in Tarski’s sense).<sup>18</sup>

The above reasoning implies that if the Gödel sentence "P" were false, it would be provable in Peano Arithmetic (hereafter “PA”), and we would then have a false theorem of PA. But, Steiner argues, this cannot be, because we have another mathematical theorem that all theorems of PA are true (in Tarski’s sense). In other words, we may safely assume that PA is sound, and this implies that there are true but unprovable sentences of arithmetic.<sup>19</sup>

There are several things to note about this interpretation of Wittgenstein’s notorious remarks on Gödel. Most importantly, it avoids engaging with the issue of  $\omega$ -inconsistency altogether--something neither Gödel’s original paper nor Wittgenstein’s remarks on that paper (as we read them) do. For it uses the standard contemporary exposition of the Gödel theorem using an improved version of Gödel’s proof due to Rosser.<sup>20</sup> Gödel showed that if PM is consistent, then P is not provable, but he did not see how to show that if PM is consistent, then  $\neg P$  is not provable. Gödel realized that the latter argument requires the stronger assumption that PM is  $\omega$ -consistent—so his proof was not “symmetric” about P and  $\neg P$ . In 1936 (in a paper that so far as we know Wittgenstein never saw) Rosser showed that this asymmetry, and Gödel’s stronger

---

<sup>18</sup> Steiner, “Wittgenstein as his Own Worst Enemy”, p. 267.

<sup>19</sup> Compare Bays, p. 200, for a similar argument; in his posted manuscript “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.” he says that he did not intend to represent Wittgenstein’s, Steiner’s, or our argument here, but only an accessible version of the Gödel proof.

<sup>20</sup> So did Steiner, “Wittgenstein as his Own Worst Enemy”, pp. 259, 262.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

assumption of  $\omega$ -consistency, are not necessary for a proof of the incompleteness of PM.<sup>21</sup> For we can construct a sentence of PM, call it “R”, such that a proof of R in PM can be converted algorithmically into a proof of  $\neg R$ , *and vice versa*, assuming only the simple consistency of PM. Freely translated, “R” expresses “for all n, if n is the Gödel number of a proof of R, then there is a proof of  $\neg R$  with a Gödel number smaller than n”, or, even more freely, “If R is provable, then  $\neg R$  is provable even more quickly”.

It strikes us as important that “R” was not the sentence “P” which either Wittgenstein or our “Note” were discussing! Steiner’s shift to a new context—one that dispenses altogether with the sentence of PM Wittgenstein and Gödel were contemplating—leaves behind the whole issue of a sentence that *purports* to say of itself that it is true but unprovable, *full stop*. Now for Steiner’s purposes, this doesn’t really matter: his reading, whether correct or incorrect, applies irrespective of whether we are considering Rosser’s R or Gödel’s P. But since P was the sentence and  $\omega$ -inconsistency the issues in which Gödel and Wittgenstein were, as we see it, interested, Steiner’s reading fails to engage with the philosophical claim we take Wittgenstein to have been investigating in his remark. If one assumes (as Steiner does) that Wittgenstein would not have minded if one replaces the Gödel sentence P by the Rosser sentence R, then the perfectly correct observation that if the undecidable sentence were refuted, then PM would have no model in which the predicate that has been interpreted as “x is a natural number” possesses an extension which is isomorphic to the natural numbers, would be *trivial*. For if R is refutable, then PM would be inconsistent and have no models at all, whereas if the Gödel sentence P is refuted PM could still be consistent.<sup>22</sup>

Here is Steiner’s pithy summary of Gödel’s argument; “[Gödel’s theorem] exhibited a ‘computer program’ which converts a proof of P into a proof of ‘not-P’ and *vice versa*.” (Steiner 2001, 262). In a note (262, n. 19) he adds “It is worth repeating that the ‘vice versa’ is due to Rosser not Gödel.” But one should be careful here about Steiner’s “vice versa”. As Steiner himself repeatedly points out in his paper, this was *not*

---

<sup>21</sup> J.B. Rosser, “Extensions of Some Theorems of Gödel and Church, *The Journal of Symbolic Logic* 1 (1936): 87-91.

<sup>22</sup> Bays agrees that the move to the context of the Rosser sentence would “seriously trivialize” our discussion and would be “quite unfair” to it; see “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.”, p. 3.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

Gödel’s argument. The formula  $P$  Gödel constructed is such that Gödel indeed showed us how to exhibit a ‘computer program’ which converts a proof of it into a proof of ‘not- $P$ ’, and that is how Gödel proved that if PM is consistent, then  $P$  is not provable. But Gödel did not prove “vice versa”, that is, he did not exhibit a computer program which converts a proof of ‘not- $P$ ’ into a proof of  $P$ . In fact, Warren Goldfarb has pointed out that if PM is  $\omega$ -consistent, then it cannot be proved even in a system as rich as PM that there is such an algorithm!<sup>23</sup> (By contrast, Gödel’s own proofs can, as is well known, be carried out in finitistic arithmetic.) To repeat: Rosser did indeed show how to construct a sentence (called “ $R$ ”) such that a proof of  $R$  can be converted into a proof of not- $R$  and vice versa. Thus Rosser showed that *if PM is consistent, then neither  $R$  (the “Rosser sentence”) nor its negation is provable*. But Gödel did not see how to show the existence of undecidable sentences in PM without a stronger hypothesis than simple consistency. (And if he had, Wittgenstein’s paragraph would have indeed been vulnerable to Steiner’s interpretation.) Instead Gödel used the hypothesis of “ $\omega$ -consistency”, and this is what our Note on the Gödel paragraph turned on. The proof of the famous 1<sup>st</sup> incompleteness theorem (the one Wittgenstein was discussing in the notorious remarks) does not claim to show that if  $\neg P$  [“not- $P$ ”] is provable then PM is inconsistent, but only that if  $\neg P$  is provable then PM is  $\omega$ -inconsistent. (The other half of the proof, that if  $P$  is provable then PM is inconsistent, does not need the notion of  $\omega$ -consistency, and does proceed as Steiner describes.)

We claimed that what interested Wittgenstein in the section we are discussing was: What would it mean if Gödel’s undecidable proposition “ $P$ ” were actually *refuted* in PM [*Principia Mathematica*]? What Wittgenstein observed, we believe, was that if this happened then PM would have *no model in which the predicate that has been interpreted*

---

<sup>23</sup> Here is Goldfarb’s proof:

Suppose it is a theorem of PM that  $\text{Provable}(\neg P) \Rightarrow \text{Provable}(P)$ . Since  $P$  is provably equivalent to  $\neg \text{Provable}(P)$ , it must also be a theorem of PM that  $\text{Provable}(\neg \neg \text{Provable}(P)) \Rightarrow \text{Provable}(P)$ , or equivalently, that  $\text{Provable}(\text{Provable}(P)) \Rightarrow \text{Provable}(P)$ . But Löb’s Theorem states that for any formula  $F$ , if it is provable that  $\text{Provable}(F) \Rightarrow F$ , then  $F$  itself is provable; taking  $F$  to be “ $\text{Provable}(P)$ ”, we conclude that  $\text{Provable}(P)$  is provable in PM. Since “ $P$ ” is provably equivalent to “ $\neg \text{Provable}(P)$ ”, it follows that  $\neg P$  is provable in PM. Then, by Gödel’s argument, PM is  $\omega$ -inconsistent.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

as “ $x$  is a natural number” possesses an extension which is isomorphic to the natural numbers. (To use Gödel’s term, PM would be  $\omega$ -inconsistent.) And in that case, Wittgenstein claimed, one “will now presumably” give up the ‘translation’ of  $P$  by the English sentence “ $P$  is not provable”, i.e., will rethink or revise it. To repeat: we did not suggest, and we see no reason to believe, that this was intended as a *refutation* of the Gödel theorem. In fact, Gödel himself specifies the formal predicates he uses to construct his undecidable sentence by appealing only to derivability considerations—something which seems to reinforce the idea that an  $\omega$ -inconsistency in PM would lead us to question whether we have the right to translate the formal predicates by words such as “is a proof of”, and so on.

But is it credible that Wittgenstein was *that* sophisticated? The answer we gave is that we have testimony (and not from particularly sympathetic sources) that Wittgenstein thought about what are now called nonstandard models of the natural numbers, and connected them with the Gödel theorem. In discussion with Alister Watson and Turing in the summer of 1937 references were made to the issue of  $\omega$ -inconsistency; in fact Watson later credited Wittgenstein with his understanding of it.<sup>24</sup> And in an essay by Goodstein from 1957 we find this:

Wittgenstein with remarkable insight said in the early thirties that Gödel’s results showed that the notion of a finite cardinal could not be expressed in an axiomatic system and that formal number variables must necessarily take values other than natural numbers; a view which, following Skolem’s 1934 publication, of which Wittgenstein was unaware, is now generally accepted.<sup>25</sup>

---

<sup>24</sup> In our “Note” (pp. 627ff) we discussed Watson’s “Mathematics and Its Foundations”, *Mind* XLVII (1938): 440-451; see also Floyd, “Prose vs. Proof: Wittgenstein on Gödel, Tarski and Truth” where discussion of the Watson essay occurred. Watson’s paper makes clear that he discussed the relevant undecidability results and the foundations of mathematics with Wittgenstein and Turing in the summer of 1937, just before Wittgenstein, having travelled to Norway in September, wrote down the notorious remarks. The parallels between the structure of the notorious remarks and the Watson paper’s presentation of Gödel are, we believe, no accident.

<sup>25</sup> Goodstein, R.L., “Critical Notice of *Remarks on the Foundations of Mathematics*” *Mind* 1957: 549-553; quotation from p. 451. (Goodstein here seems not to understand the Skolem result, for it is about true arithmetic, not about any axiomatized system.) In another essay (“Wittgenstein’s Philosophy of Mathematics”, in A. Ambrose and M. Lazerowitz eds., *Ludwig Wittgenstein: Philosophy and Language* (London: Allen & Unwin, 1972), pp. 271-286) Goodstein says (p. 279): “I [Goodstein] do not think Wittgenstein heard of Gödel’s discovery before 1935; on hearing about it his immediate reaction, with I think truly remarkable insight, was to observe that it showed that the formalization of arithmetic with mathematical induction and the substitution of numerals for variables fails to capture the concept of natural number, and the variables must admit values which are not natural numbers. For if, in a system  $A$ , all the

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

If this *was* Wittgenstein’s interest in the Gödel paragraph, then it was absolutely necessary that he should consider the possibility that PM was *unsound* (since  $\omega$ -inconsistency counts as unsoundness, for a system which is supposed to formalize mathematics). By contrast, there is no known reason to think that Wittgenstein was interested in the possibility that *Peano Arithmetic* is  $\omega$ -inconsistent; Peano arithmetic is unquestionably sound mathematics, while *Principia Mathematica* was, we believe “set theory” (or as bad as set theory) in Wittgenstein’s eyes, and thus essentially a misleading construal of mathematics, if not metaphysics in formal dress.

These issues are important because, as our reconstruction of his argument makes clear, Steiner also assumes that Wittgenstein would not have minded if one replaced “PM” in the Gödel paragraph by “PA” (Peano Arithmetic).<sup>26</sup> Thus although what Steiner writes about *PA* is true, it is *not* the case that we have “another *mathematical theorem*” that all theorems of *PM* are true (in Tarski’s sense) in any system which is not essentially stronger than PM itself – any system which is not a rich system of what we would all regard as “set theory”. And since Wittgenstein’s remark was about PM and not about PA, the applicability of Tarski’s theory of truth to PA is irrelevant. Steiner’s assumption is that the soundness of PA can be proved in whatever system we employ to formalize Tarski’s theory of truth. But that is not the same as assuming that that system can prove the soundness of Russell’s *Principia Mathematica*! If we are right in our understanding of what Wittgenstein has in mind in writing “suppose the opposite is proved in Russell’s system”, then his whole point was to ask whether we would hold on to our English interpretation of P as “P is not provable” if  $\neg P$  were proved *and we therefore realized that PM was not sound*. There is no reason to think that Wittgenstein would have regarded PA in the same light as PM—and no reason to think that we, looking back all these decades later, should do so either.<sup>27</sup>

---

sentences  $G(n)$  with  $n$  a natural number are provable, but the universal sentence  $(\forall n)G(n)$  is not, then there must be an interpretation of  $\mathcal{A}$  in which  $n$  takes values other than natural numbers for which  $G(n)$  is not true (in fact in 1934, Th. Skolem had shown that this was the case, independently of Gödel’s work).“

<sup>26</sup> The first footnote to (Steiner 2001) includes the sentence: “For the purpose of this essay, “Russell’s system” can be understood as first-order Peano arithmetic.” Bays too offers this substitution, as we shall see below.

<sup>27</sup> With this Steiner and Bays would presumably agree; see Bays 2004 207 and his “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.”

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

This is not merely an historical aside, it is a logical point. Steiner’s criticism of the argument he attributes to Wittgenstein ultimately turns on invoking a semantical proof of the Gödel theorem that is a non-constructive proof in set theory. Tarski himself did not actually give such a proof, but his definition of *satisfaction-in-a-model* does yield the means to express in set theory, not only the notion of a sentence being *true in arithmetic*, but also the idea of a set being first-order *definable* in a formalized language of arithmetic.<sup>28</sup> Putting this together with Gödel’s manner of defining the notion of *provable in arithmetic* (via Gödel numbering), we can then express the following argument in set theory:

*Truth in arithmetic* is not definable in a formalized theory of arithmetic, but, as Gödel showed, *provability in arithmetic* is. Hence the two notions cannot be coextensive.

Gödel claimed to have already grasped this (non-constructive) argument in 1930, but he did not publish it, partly because of the finitistic context of the Hilbert Program, and partly out of worry that philosophers hostile to the notion of *truth* might object to his result.<sup>29</sup> What Gödel’s 1931 paper (the one Wittgenstein saw!) gives, quite self-consciously, is a different, finitistically acceptable argument for the incompleteness of arithmetic *via* an explicit construction of an undecidable sentence. As Gödel wrote to Menger after reading Wittgenstein’s notorious remarks about his theorem, his incompleteness result “is a mathematical theorem within an absolutely uncontroversial part of mathematics (finitary number theory or combinatorics)”.<sup>30</sup>

---

<sup>28</sup> For a first-order language  $L$ , a structure  $\mathbf{M}$  interpreting  $L$ , a formula  $\psi$  of  $L$  with  $k$  free variables and  $a_1, \dots, a_k$  elements of the domain  $M$  of  $\mathbf{M}$ , “ $\mathbf{M} \models \psi [a_1, \dots, a_k]$ ” is taken to assert that the formula  $\psi$  is satisfied in  $\mathbf{M}$  when the  $k$  free variables are assigned sequentially to  $a_1, \dots, a_k$ . Tarski’s definition of satisfaction in a structure (or model), his analysis of “truth in an interpretation”, amounts to a precise, set-theoretic definition of this notion. Applied to the particular structure of arithmetic that model theorists call “ $\mathbf{N}$ ” (to be characterized in the final section of our essay, below) it provides an analysis of the notion of a sentence being *true in arithmetic*. Now with  $L$  and  $\mathbf{M}$  as above, a set  $X \subseteq M$  is *definable in  $L$*  if there is a formula  $\psi$  in  $n+1$  free variables and “parameters”  $b_1, \dots, b_n$  such that  $X = \{a \in M \mid \mathbf{M} \models \psi [a, b_1, \dots, b_n]\}$ .

<sup>29</sup> See Hao Wang, *Reflections on Kurt Gödel* (Cambridge, MA., MIT Press, 1987) pp. 84ff.; cf. J.W. Dawson, Jr., *Logical Dilemmas: The Life and Work of Kurt Gödel* (A.K. Peters: Wellesley, MA, 1997), chapter IV.

<sup>30</sup> *Kurt Gödel Collected Works Vol. V, Correspondence H-Z*, eds. S. Feferman et.al. (New York: Oxford University Press, 2003), p. 133.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

If we assume that arithmetic is bivalent<sup>31</sup> and look at the Gödel theorem not through the eyes of Rosser’s improvement, but with (what we are taking to be) Wittgenstein’s point in mind, then we can say that what the Gödel theorem shows is *either* that (1) there exists a true but unprovable sentence of number theory in PM *or* that (2) number theory is inexpressible in PM. For if PM is consistent, and yet we had a proof that  $\neg P$ , then Gödel’s theorem tells us that PM is  $\omega$ -inconsistent. And then there will be no way of defining in the model-theorist’s sense (see footnote 28) the set of natural numbers *in* PM.

### 3. Bays’s Reply to our Note

According to Bays, the philosophical upshot of our interpretation is unacceptable in two ways. 1) It ultimately urges upon readers the abandonment of  $\mathcal{N}$ , the standard model of arithmetic, because the “core” of our argument (according to Bays) is that “what a given formula ‘expresses’ depends on the model at which we interpret it” (Bays, pp. 201-2, p. 204), and we give insufficient weight to what Bays repeatedly calls the “canonicity” of  $\mathcal{N}$ . 2) It forces us to deny the meaningfulness of what Bays takes to be clearly meaningful if not true, namely, the claim that there *are* true but unprovable propositions of arithmetic. The sense of this claim’s meaningfulness is explored at length by Bays, primarily through a series of counterfactuals: he argues that if Gödel’s  $\neg P$  were derived in PA (or PM), mathematicians would most likely hunt for new axioms in light of non-standard results about a particular formalized theory, and they would certainly not abandon  $\mathcal{N}$ ; and this hypothetical “abandonment” he takes us to be committed to “urging” (Bays, pp. 201ff). But we never did discuss these counterfactuals, and in fact neither 1) nor 2) follows from our reading of Wittgenstein’s notorious remarks.

In his paper Bays takes for granted certain views about what Wittgenstein is up to; though recently he has pulled back from endorsing these as readings of Wittgenstein, his apparent presumptions are widely enough shared that they are worth discussing in some

---

<sup>31</sup> The assumption that arithmetic is bivalent can itself be given two very different interpretations: it can mean either 1) that the schema  $p \vee \neg p$  is accepted as part of the deductive apparatus of arithmetic, or 2) that every sentence of arithmetic is (in language introduced by Dummett, who repeatedly emphasizes the difference between these assumptions) that every sentence of arithmetic is “determinately true or false” (see Michael Dummett, *The Logical Basis of Metaphysics* (Cambridge, MA.: Harvard University Press, 1991), pp. 74ff. Here we have in mind the stronger assumption.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

detail.<sup>32</sup> (If Bays’s own harsh criticisms of our arguments applied, they would presumably also impugn our argument as reasonable interpretation of Wittgenstein.) First, Bays states that “Wittgenstein is criticizing a relatively common interpretation of Gödel’s first theorem: that the theorem shows—or helps to show—that there are true but unprovable sentences of ordinary number theory” (Bays 2004 198). What is relevant here, however, is the difference between our view of this critique and Mark Steiner’s, the latter being the most sophisticated version yet published of the (received) view that Wittgenstein (perhaps unwittingly) is presenting a line of thought the conclusion of which denies something that is true, viz., “there are true but unprovable sentences of ordinary number theory”. Like Steiner, Bays thinks that a) Wittgenstein objects to the “common interpretation” of the Gödel theorem “partially because he is skeptical concerning the notion of “truth” in play here (‘True’ in what system?)” (Bays 2004 198); he adds that b) Wittgenstein is also “opposed in principle to the derivation of ‘philosophical’ claims from ‘mathematical’ arguments” (Bays 2004 198; Steiner has made no such claim). Our interpretation, by contrast, takes Wittgenstein to have been attempting a constructive clarification of the idea of “‘true’ in a system” by way of his understanding of the actual details of Gödel’s 1931 proof: his ‘skepticism’ concerns (as we see it) claims of a specific interlocutor (perhaps himself!) not a general contextualism or skepticism about “Truth” (or, e.g., bivalence). We believe that no such *general* principle as b) is to be found in Wittgenstein’s later thought.

At issue is of course not merely Wittgenstein interpretation, but the appropriate response to certain philosophical claims about Gödelian incompleteness. Bays correctly sees that our reasoning does not entail that it is *false simpliciter* to say that “Gödel’s incompleteness theorem shows that there are true but unprovable sentences of number theory” (p.203), but rather that it is *unclear* until we look at the details of the proof. When we do, the original translation of P by the English sentence ‘P is not provable’ must be “given up” in the situation supposed. What is worthy of discussion is precisely what is meant by “giving up” the original translation. Wittgenstein makes it clear that he

---

<sup>32</sup> In his subsequent manuscript “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.”, Bays asserts that he really did not wish to take any stand on interpreting Wittgenstein. Hence he is withdrawing the statements about Wittgenstein (in his 2004), as well as his apparent acceptance of Steiner’s interpretation of the notorious remarks at Bays 2004, 208, n. 27.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

is concerned with the translation into *German*, i.e., the translation into ordinary language. This, as we see it, is crucial for determining the aim and structure of Wittgenstein’s philosophical *investigation* of the Gödel proof.

Like Steiner, Bays assumes that the substitution of PA for PM in reading Wittgenstein’s notorious remarks will not affect our understanding of their philosophical significance. He writes,

Wittgenstein focuses on number theory as formulated in “Russell’s system” – that is, the system of *Principia Mathematica*. There is, however, nothing in his argument that depends on this particular choice of background logic. For expository convenience I will recast the argument in terms of ordinary, first-order Peano Arithmetic. Later...I discuss the possible philosophical significance of formulating the argument in “Russell’s system” (Bays 2004, 198 n. 4).

In our rendition of Wittgenstein’s notorious paragraph, the force of his insight crucially depends on the fact that he is discussing PM. It is true that in Wittgenstein’s day, PM was taken by many to formalize *all* of mathematics (though not, of course, by Wittgenstein). But when Bays offers what he calls a “sketch of the claims Floyd and Putnam want to argue against” and an account of the “core” of our argument, he presents these as a matter of imagining that PA, and not PM, might turn out to be unsound (cf. Bays 2004, 200-201).<sup>33</sup> He admits that given the relative security of the axioms of PA, the former scenario is “extremely implausible” and “would require such deep revisions of present mathematics that it is virtually impossible to adjudicate questions concerning ‘what we would/should do’ in such circumstances” (Bays 2004 205). With this we would agree. But Bays goes on to so adjudicate, and states that our argument “surely” entails only one answer to the “implausible” counterfactual, something that is for Bays “*unimaginable*”, viz., that in this “implausible” scenario mathematicians would “adopt recognizably nonstandard models of arithmetic as canonical for interpreting the language of number theory” (Bays 2004 204-5) (and thereby throw out PA (or PM)).

We of course never discussed the “implausible” counterfactual at all, and it is difficult to see how our interpretation would entail any position with respect to it. But the

---

<sup>33</sup> In “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.” he continues to say that his argument “follows Floyd and Putnam in supposing that we have discovered a proof that PA is *not* sound” (p. 8). But again: we never discussed PA, and do not take Wittgenstein to have been doing so.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

consequences of Bays’s shift in presenting what we assumed are significant for his reading of our ‘Note’. For our interpretation, according to Bays, entails that it is impermissible for anyone to “step back to make semantic generalizations”, or to “recognize from the *outside* that [e.g.] contradictions are unacceptable” (Bays 2004 210 n. 32). As Bays insists (presumably with us and/or Wittgenstein in mind), “we certainly have *some* ability to step back and engage in critical reflection on (purported) foundations” (Bays 2004 210 n. 32).

Of course we have such ability, and our reading of the notorious remarks, whether wrong or right about Wittgenstein, shows one instance of critical reflection at work. But Bays claims to find in our discussion the suggestion that, were PA to turn out to be unsound, “only models which satisfy PA should count as ‘admissible interpretations’ for our language” (2004, 204), and this leads him to (erroneously) assert that “the insistence that we limit ourselves to models of PA when we interpret arithmetic runs rather deep in Floyd and Putnam’s paper” (2004 204 n. 17).<sup>34</sup> In the end, then, Bays thinks that our reading of the notorious remarks must work against conceiving the standard model of arithmetic,  $\mathbf{N}$ ,  $\mathbf{Th}(\mathbf{N})$  -- and even model theory generally -- as viable objects of study for mathematicians. He alleges that our “Note” leads ineluctably to “the abandonment of  $\mathbf{N}$  that Floyd and Putnam urge upon us” (2004 204, n. 19), and suggests that this could even lead to a principled rejection of “the semantical analysis of formal systems”, to a “crazy” position which is offering “principled reasons for permitting us to step back to make syntactic generalizations, while forbidding us to step back to make semantic generalizations” (2004 210, cf. n 31).<sup>35</sup>

The reader will see no trace of these ideas, either in our “Note” or in the argument sketched above. So what has happened?

Although Bays says he will make no interpretative claims about Wittgenstein (Bays 2004, 197), he makes several suggestions about the notorious remarks and our

---

<sup>34</sup> Referring to earlier work of Putnam’s, he holds that Putnam, in particular, is committed to “a similar insistence on the priority of axioms over interpretations” (2004 204 n. 18). But we are not applying earlier arguments of Putnam’s: see footnote 37 below.

<sup>35</sup> At 2004 210 n. 31 Bays attributes this view quite explicitly to us (or perhaps to Wittgenstein), asserting that “As a point *tu quo*, I would note that Floyd and Putnam themselves engaged in purely semantic reasoning about Russell’s system”. Of course we do so engage; this should have suggested to Bays that he had misinterpreted our position.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

interpretation of them that are, unfortunately, mistaken. Since these are often linked to Wittgenstein explicitly (without interpretative argument), they are worth examining in themselves, even if now Bays would not wish to ascribe any particular interpretation to Wittgenstein. The first is the suggestion that, inspired by Wittgenstein, we wrote about Gödel in the grip of anti-realism, verificationism, and/or formalism, and that we identified a priori the notions of ‘mathematically true’ and ‘proved’ (or perhaps, less stringently, ‘provable’ or ‘provable in some formal system or other’). As Bays writes, “at best, Floyd and Putnam suggest, Gödel’s paper gives rise to a proof-theoretic conception of truth. It does follow immediately from this identification that there can be no true but unprovable sentences of arithmetic” (Bays 2004, 202 n 14).<sup>36</sup> But, contrary to what Bays says, the argument presented in our "Note" neither depends upon nor was used to advocate such a philosophical point of view: as we have said, one of our main points was to offer a reading the notorious paragraph that contains something more than a facile philosophical prejudice identifying the notions of *truth* and *proof*.

Nor does our reading depend upon the assumption, which Bays quite explicitly attributes to Wittgenstein at the opening of his paper, that one should be “opposed in principle to the derivation of ‘philosophical’ claims from ‘mathematical’ arguments” (Bays 2004, 198). On the contrary: our interpretation credits Wittgenstein with sensitivity to the subtle interplay between ordinary and formalized language in the context of philosophical disputes about the foundations of mathematics. What we were arguing is that neither formalism nor verificationism about mathematical truth in general are at issue in Wittgenstein's 1937 remarks on Gödel.<sup>37</sup> (It is worth noting that although

---

<sup>36</sup> Though in “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.” (n. 16) Bays expresses astonishment that we record this as his suggestion, but in the same manuscript he again links our interpretation to an attempt to place “pressure” in what he calls “naïve realism” (p. 12; the whole final section of the manuscript is relevant). This may be because he links our “Note” to earlier work of Putnam’s applying a model-theoretic argument to question realism (Putnam’s work and Bays’s previously published criticisms of it are referred to in Bays 2004 204 n. 18 and in “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.” at n. 27). The argument of our “Note” is not to be read, however, as an application of Putnam’s earlier model-theoretic arguments against realism.

<sup>37</sup> Several reasons for doubting that the notorious remarks express verificationism about mathematical truth as a whole are canvassed in Floyd, "On Saying What You Really Want to Say: Wittgenstein, Gödel and the Trisection of the Angle". For the larger issue of verificationist tendencies in Wittgenstein's other later remarks, see Hilary Putnam, “Was Wittgenstein *Really* an Antirealist About Mathematics?”, in eds. Timothy G. McCarthy and Sean Stidd, *Wittgenstein in America* (Oxford; Clarendon Press, 2001), pp. 140-194.

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ›berüchtigte‹ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

he does take Wittgenstein to be identifying “true in *Principia Mathematica*” with “proved in (the game of playing with) *Principia Mathematica*” in the notorious paragraphs, Steiner himself rejects, just as we do, the idea that Wittgenstein was a verificationist.<sup>38</sup>)

In closing, we must consider Bays's reasoning about the so-called "standard" model of arithmetic and the notion of *truth*. Throughout his paper Bays assumes that the notions of *truth*, *interpretation* and *natural number* are and ought to be determined from within model theory alone, with reference—indeed, what Bays calls “perfectly canonical” reference—to the set (or model)  $\mathbf{N}$  (Bays 2004, 210).

Bays never defines  $\mathbf{N}$ , but let us do so now.  $\mathbf{N}$  is a *structure* in the model-theorist’s sense, that is, a set conceived of as coming to us equipped with two distinguished elements, zero and one, and the functions of addition and multiplication.  $\mathbf{N}$  is to be regarded as an interpretation for a corresponding first-order language  $L$  containing the usual apparatus of quantification theory, two binary function symbols, and two constant symbols. The *full theory of  $\mathbf{N}$* ,  $\mathbf{Th}(\mathbf{N})$ , sometimes called by model theorists “true arithmetic”, is defined as the set of sentences of  $L$  that are satisfied in  $\mathbf{N}$  in the sense of Tarski;  $\mathbf{Th}(\mathbf{N})$  is thus defined recursively via Tarski’s notion of *satisfaction in a structure*.

We are given by this construction that every sentence of  $L$  will either be in  $\mathbf{Th}(\mathbf{N})$  or it will not be. What this definition does not, however, tell us is how to determine, for any arbitrarily given particular sentence of  $L$ , whether that sentence is or is not an element of  $\mathbf{Th}(\mathbf{N})$ . Indeed, by establishing that  $\mathbf{Th}(\mathbf{N})$  is in principle not (recursively) axiomatizable, Gödel showed there *is* no way of effectively determining this. He thereby showed that  $\mathbf{Th}(\mathbf{N})$  can have no axiomatization--that, in other words, every axiomatization is either incomplete (and so must have non-standard models) or is  $\omega$ -inconsistent (so has only non-standard models) or is inconsistent (and has no models at all).<sup>39</sup>

---

<sup>38</sup> Personal correspondence; compare Steiner’s “Wittgenstein: Mathematics, Regularities and Rules” in *Benacerraf and His Critics*, eds. Adam Morton, Stephen Stich (Cambridge, Mass.: Blackwell Publishing Company, 1996), pp. 190-212.

<sup>39</sup> Note that Gödel’s proof applies to *any* first-order formalized language  $L^*$  sufficiently rich to express the usual arithmetical operations of addition, multiplication and quantification on the natural numbers. Thus even if it can express more than the minimal language  $L$ , any such  $L^*$  will fail to recursively axiomatize a

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

Bays tries to eliminate the issue of  $\omega$ -consistency by reiterating several times that  $\mathbf{N}$  is a “core” or “perfectly canonical” “interpretation” of “mathematical truth” applicable to every formula of PM and our natural language in the same way, across a wide variety of counterfactual contexts (cf. Bays, pp. 204, 206, 208, 210). But why Bays thinks that we should assume that the notion of *interpretation* is reducible to the model-theorist’s notion of  $\mathbf{N}$  he does not say. Nor does he explain what he means by a “perfectly canonical” “interpretation”.

We for our part were careful to distinguish, as most philosophers do, between the notions of *interpretation*, *model* and *translation*. In terms of the notorious paragraphs under discussion, the relevant contrast is between *Übersetzung* (translation) and *Deutung* (interpretation, implying clarification), a distinction which Wittgenstein’s German and our discussion of the notorious paragraph respect. Unfortunately Bays slides, equivocally, among them.<sup>40</sup> On our view, whether a given sentence of PM does or does not have a model of a particular kind, subject to a particular assumption about PM’s being  $\omega$ -consistent, affects but does not exhaust the issue of whether, why, and to what extent we can translate or interpret in an informal sense sentences of PM into English apart from that assumption. As we explicitly wrote in our “Note”, a kind of translation (a “correlation” or “transformation”) of sentences of PM into English might be possible even if it turned out that there were no models, hence no admissible interpretations, of

---

full theory of the structure consisting of the natural numbers and the functions of addition and multiplication on these numbers.

<sup>40</sup> A running together of the notions of “model” and “interpretation” occurs when Bays mistakenly says that the “core” of our argument is the idea that “*all* ‘admissible interpretations’ [of arithmetic] take place on nonstandard models [because]...what a formula ‘expresses’ depends on the model at which we interpret it” (2004 202-3). What a formula ‘expresses’ may depend upon there being a model, or it may not, as we say (“Note”, p. 626)); this just shows that we have *not* reduced the notion of interpretation (or of truth!) to something *only* model-relative. Note, for the same running together, Bays 2004 204 n.17: “the insistence that we limit ourselves to models of PA when we interpret arithmetic runs rather deep in Floyd and Putnam’s paper. At one point they even suggest that, were we to find PA inconsistent, we should conclude that there are no admissible interpretations of arithmetic (see [Floyd and Putnam, p.] 626)”. As is easily verified, we never spoke about all admissible interpretations of arithmetic (nor even about PA), but, following Wittgenstein, only about “admissible interpretations of PM” (Bays has simply misread and misquoted the pronoun “it” in the passage he cites). By “admissible interpretation” we understood an interpretation “fitting” at least one model of PM (cf. pp. 625-6). If PM (or any other formal theory) is simply inconsistent, there will be no models of it, hence *a fortiori* there will be no admissible interpretations of it (in our sense), though there could be (in our sense) translations and improvements of it. The relevant contrast is not between the merely “syntactic” and that which “involves interpretation” in the model theorist’s sense, as Bays seems to assume (cf. 2004 202); nor is there any “insistence” on prioritizing axioms over models (despite Bays 2004 204).

To appear in „Wittgensteins ‚berüchtigter‘ Paragraph über das Gödel-Theorem: Neuere Diskussionen“ (with Hilary Putnam), in *Prosa oder Beweis? Wittgensteins ‚berüchtigte‘ Bemerkungen zu Gödel, Texte und Dokumente*, Esther Ramharter hrsg., (Berlin: Parerga Verlag, 2008), pp. 75-97.

PM itself as a whole (Floyd and Putnam, p. 626). In short, the relations among the notions of *model*, *interpretation*, and *translation* are as complex as the relations of any of these notions to that of *meaning*: to assess them, we need to be clear about the context within which we are using them. Wittgenstein’s insight, as we understand it, is but a special case of this more general point.

To summarize: Bays assumes throughout his paper that his remarks have a direct bearing on philosophical problems about the notion of *truth*. Yet we find it odd that Bays finds it reasonable to grant (1) that there exist philosophical problems about arithmetical truth, and also (2) that the existence of model theory will resolve these problems. In fact, Bays seems to think that an appeal to  $\mathbf{N}$  proves, not only that the Gödel sentence has a truth-value, but that every sentence of arithmetic has a truth-value—that, in other words, the principle of bivalence holds in arithmetic.<sup>41</sup>

It is not that the authors of the “Note” are defenders of any philosophy on which bivalence fails in arithmetic.<sup>42</sup> But neither would we want to read intuitionists (Brouwer, Heyting) or nominalists, finitists, and so on out of the camp of philosophy. And we certainly would not claim that Brouwer, Goodman, etc. have been *refuted* by Tarski, Abraham Robinson, and the other founders of model theory.

Does Bays think otherwise?

---

<sup>41</sup> In his manuscript “Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.” n. 27 Bays says that he finds our distinctions between interpretation, translation and model “quite misleading”, and our emphasis on “more nuanced and complicated” aspects of meaning irrelevant to the main “disagreement” between us which is, he speculates, the issue of (what he calls) “naïve realism” (see his section 4. “What’s the Issue?”). No such form of realism was discussed in Bays’s 2004, but in any case our “Note” was not preoccupied with any such view (see note 37 above for references to articles where realism is discussed; compare footnote 36). A discussion of Bays’s own ideas about “naïve realism” lies outside the scope of this essay.

<sup>42</sup> See Putnam’s Tarski lectures, “Paradox Revisited I: Truth”, and Paradox Revisited II: Sets—A Case of All or None?”, in *Between Logic and Intuition: Essays in Honor of Charles Parsons*, G. Sher and R. Tieszen, eds. (New York: Cambridge University Press, 2000), pp. 3-26.